Figure 1

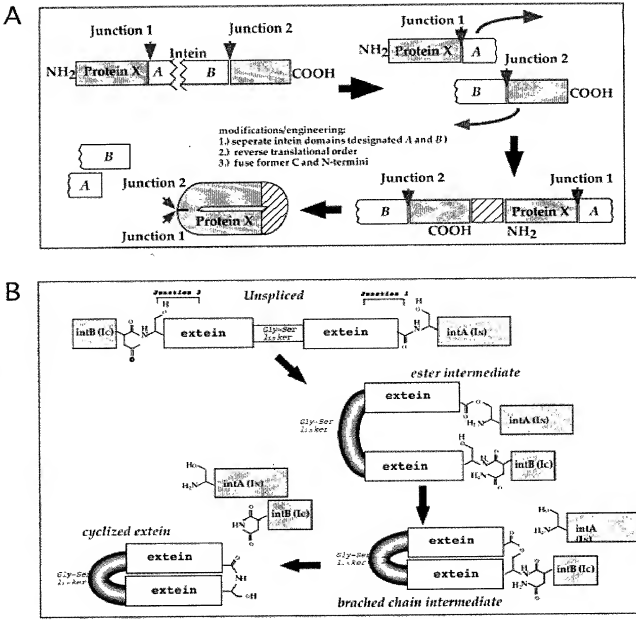
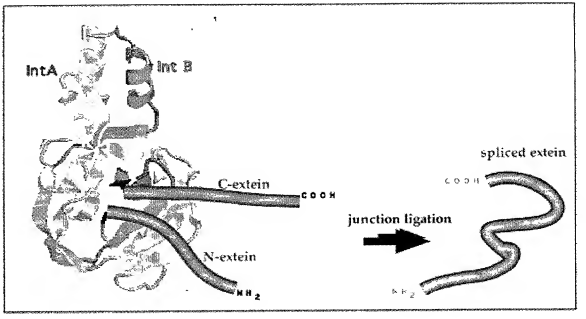

A
Junction 1
Junction 2
Intein
NH2
Protein X
A
B
COOH
modifications/engineering:
1.) seperate intein domains (designated A and B)
2.) reverse translational order
3.) fuse former C and N-termini
B
Unspliced
intB (Ic)
extein
intA (Is)
ester intermediate
cyclized extein
brached chain intermediate

Figure 2

A


IntA
Int B
C-extein
COOH
N-extein
NH2
junction ligation
spliced extein

B

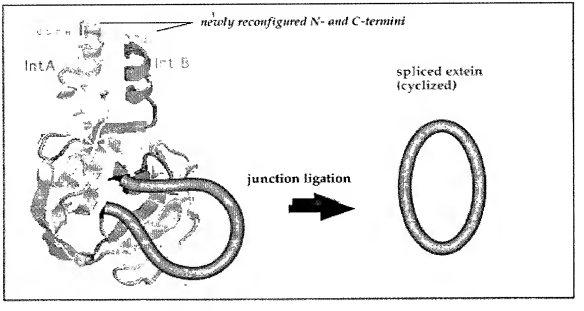

newly reconfigured N- and C-termini
IntA
Int B
junction ligation
spliced extein
(cyclized)

FIGURE 3

A)
GCISGDSLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYI
LKTRLGRTIKA
TANHRFLTIDGWKRLDELSLKEHIALPRKLESSSLQLMSDEELGLLGHLIGDGCTLPR
HAIQYTSNKIEL
AEKVVELAKAVFGDQINPRISQERQWYQVYIPASYRLTHNKKNPITKWLENLDVFGL
RSYEKFVPNQVFE
QPQRAIAIFLRHLWSTDGCVKLIVEKSSRPVAYYATSSEKLAKDVQSLLLKLGINARL
SKISQNGKGRDN
YHVTITGQADLQIFVDQIGAVDKDKQASVEEIKTHIAQHQANTNRDVIPKQIWKTYV
LPQIQIKGITTRD
LQMRLGNAYCGTALYKHNLSRERAAKIATITQSPEIEKLSQSDIYWDSIVSITETGVEE
VFDLTVPGPHN
FVANDIIVHNS

B)
YCITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHP
VYTVRTVEGLRV
TGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAP
TTYTVGVPGLVRF
LEAHHRDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFV
SHNT

C)
ECLTSDHTVLTTRGWIPIADVTLDDKVAVLDNNTGEMSYQNPQKVHKYDYEGPMY
EVKTAGVDLFVTPNH
RMYVNTTNNTTNQNYNLVEASSIFGKKVRYKNDAIWNKTDYQFILPETATLTGHTN
KISSTPAIQPEMNA
WLTFFGLWIANGHTTKIAEKTAENNQQKQRYKVILTQVKEDVCDIIEQTLNKLGFNFI
RSGKDYTIENKQ
LWSYLNPFDNGALNKYLPDWVWELSSQQCKILLNSLCLGNCLFTKNDDTLHYFSTS
ERFANDVSRLALHA
GTTSTIQLEAAPSNLYDTIIGLPVEVNTTLWRVIINQSSFYSYSTDKSSALNLSNNVAC
YVNAQSALTLE
QNSQKINKNTLVLTKNNVKSQTMHSQRAERVDTALLTQKELDNSLNHEILINKNPGT
SQLECVVNPEVNN
TSTNDRFVYYKGPVYCLTGPNNVFYVQRNGKAVWTGNS

FIGURE 3

D)
LCVAPETMILTEDGQFPIKDLEGKIIKVWNGNEFSSVTVVKTGTEKELLEVELSNGCT
LSCTPEHKFIIV
KSYTEAKKQKTDDNAIANAERVDAQDLKPRMKLIKFDLPTLFGNSEHDIKYPYTHGF
FCGDGTYTKYGKP
QLSLYGDKKELLTYLDVRTMTGLEDASGRLNTWLPLDLAPKFDVPINSSLECRMEW
LAGYLDADGCVFRN
GTNESIQVSCIHLDFLKRIQLLLIGMGVTSKITKLHDEKITTMPDGKGGQKPYSCKPIW
RLFISSSGLYH
LSEQGFETRRLKWEPRQPQRNAERFVEVLKVNKTGRVDDTYCFTEPINHAGVFNGIL
TGQC

E)
GCFTKGTQVMMADGADKSIESIEVGDKVMGKDGMPREVVGLPRGYDDMYKVRQL
SSTRRNAKSEGLMDFT
VSADHKLILKTKQDVKIATRKIGGNTYTGVTFYVLEKTKTGIELVKAKTKVFGHHIH
GQNGAEEKAATFA
AGIDSKEYIDWIIEARDYVQVDEIVKTSTTQMINPVHFESGKLGNWLHEHKQNKSLA
PQLGYLLGTWAGI
GNVKSSAFTMNSKDDVKLATRIMNYSSKLGMTCSSTESGELNVAENEEEFFNNLGA
EKDEAGDFTFDEFT
DAMDELTINVHGAAASKKNNLLWNALKSLGFRAKSTDIVKSIPQHIAVDDIVVRESLI
AGLVDAAGNVET
KSNGSIEAVVRTSFRHVARGLVKIAHSLGIESSINIKDTHIDAAGVRQEFACIVNLTGA
PLAGVLSKCAL
ARNQTPVVKFTRDPVLFNFDLIKSAKENYYGITLAEETDHQFLLSNMALVHNC

F)
GCLSYATNQPYFLKSDNVNFSKLTSLKVSNHYILSATLELLIPFQYNRIYPIVSLIKREL
QTGYKVVYEL
DFYISVIVSTVEHYVLTLNGWKRILELTVDDLVATLDIQYLIYNNTEVDLFSSNVIFSS
VINLICMNRIN
VYDFWIPKTNNFFVNALLVHNS

G)
GCISKFSHIMWSHVSKPLFNFSIKKSHMHNFNKNIYQLLDQGEAFISRQDKKTTYKIR
TNSEKYLELTSN
HKILTLRGWQRCDQLLCNDMITTQIGFELSRKKKYLLNCIPFSLCNFETLANINISNFQ
NVFDFAANPIP
NFIANNIIVHNS

FIGURE 3

H)
GCFAKGTNVLMADGSIECIENIEVGNKVMGKDGRPREVIKLPRGRETMYSVVQKSQ
HRAHKSDSSREVPE
LLKFTCNATHELVVRTPRSVRRLSRTIKGVEYFEVITFEMGQKKAPDGRIVELVKEVS
KSYPISEGPERA
NELVESYRKASNKAYFEWTIEARDLSLLGSHVRKATYQTYAPILYENDHFFDYMQK
SKFHLTIEGPKVLA
YLLGLWIGDGLSDRATFSVDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVN
LYSKVVRGNGIRNN
LNTENPLWDAIVGLGFLKDGVKNIPSFLSTDNIGTRETFLAGLIDSDGYVTDEHGIKA
TIKTIHTSVRDG
LVSLARSLGLVVSVNAEPAKVDMNGTKHKISYAIYMSGGDVLLNVLSKCAGSKKFR
PAPAAAFARECRGF
YFELQELKEDDYYGITLSDDSDHQFLLANQVVVHNC

I)
GCFAYGTRGALADGTTEKIGKIVNQKMDVEVMSYDPDTDQVVPRKVVNWFNNGPA
EQFLQFTVEKSGGNG
KSQFAATPNHLIRTPAGWTEAGDLVAGDRVMAAEPHRLSDQQFQVVLGSLMGDGN
LSPNRRDRNGVRFRM
GHGAKQVDYLQWKTALLGNIKHSTHVNDKGATFVDFTPLPELAELQRAVYLGDGK
KFLSEENFKALTPLA
LVFWYMDDGPFTVRSKGLQERTAGGSGRIEICVEAMSEGNRIRLRDYLRDTHGLDV
RLRLSGAAGKSVLV
FSTASSAKFQELVAPYITPSMEYKLLPRFRGQGAVTPQFVEPTQRLVPARVLDVHVK
PHTRSMNRFDIEV
EGNHNYFVDGVMVHNS

J)
YCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEVLE
YELEDGSVIR
ATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLDAGTIK

K)
KALALDTPLPTPTGWTAMGDVAVGDELLAVDEAPTRVVAATEVMLGRPCYEIEFSD
GTVIVADAQHQWPT
SYGIRTSAQLRCGLDIIAAAGSTPRHAGRLTTAAFMAPVLCIDSVRRVRSVPVRCVEV
DNAAHLYLAGRG
MVPTHNS

FIGURE 3

L)

GALAYDEPIYLSDGNIINIGEFVDKFFKKYKNSIKKEDNGFGWIDIGNENIYIKSFNKLS
LIIEDKRILR
VWRKKYSGKLIKITTKNRREITLTHDHPVYISKTGEVLEINAEMVKVGDYIYIPKNNTI
NLDEVIKVETV
DYNGHIYDLTVEDNHTYIAGKNEGFAVSNC

M)

GALYDFSVIQLSNGRFVLIGDLVEELFKKYAEKIKTYKDLEYIELNEEDRFEVVSVSP
DIKANKHVVS
RVWRRKVREGEKLIRIKTRTGNEIILTRNHPLFAFSNGDVVRKEAEKLKVGDRVAVM
MRPPSPPQTKA
VVDPAIYVKISDYYLVPNGKGMIKVPNDGIPPEKAQYLLSVNSYPVKLVREVDEKLS
YLAGVILGDGY
ISSNGYYISATFDDEAYMDAFVSVVSDFIPNYVPSIRKNGDYTIVTVGSKIFAEMLSRI
FGIPRGRKS
MWDIPDVVLSNDDLMRYFIAGLFDADGYVDENGPSIVLVTKSETVARKIWYVLQRIG
IISTVSRVKSR
GFKEGELFRVIISGVEDLAKFAKFIPLRHSRKRAKLMEILRTKKPYRGRRTYRVPISSD
MIAPLRQML
GLTVAELSKLASYYAGEKVSESLIRHIEKGRVKEIRRSTLKGIALALQQIAKDVGNEE
AWVRAKRLQI
IAEGDVYWDEVVSVEEVDPKELGIEYVYDLTVEDDHNYVANGILVSNC

N)

PCVSGDTIVMTSGGPRTVAELEGKPFTALIRGSGYPCPSGFFRTCERDVYDLRTREGH
CLRLTHDHRVL
VMDGGLEWRAAGELERGDRLVMDDAAGEFPALATFRGLRGAGRQDVYDATVYGA
SAFTANGFIVHNC

O)

GCIDGKAKIIFENEGEEHLTTMEEMYERYKHLGEFYDEEYNRWGIDVSNVPIYVKSF
DPESKRVVKGKVN
VIWKYELGKDVTKYEIITNKGTKILTSPWHPFFVLTPDFKIVEKRADELKEGDILIGGM
PDGEDYKFIFD
YWLAGFIAGDGCFDKYHSHVKGHEYIYDRLRIYDYRIETFEIINDYLEKTFGRKYSIQ
KDRNIYYIDIKA
RNITSHYLKLLEGIDNGIPPQILKEGKNAVLSFIAGLFDAEGHVSNKPGIELGMVNKRL
IEDVTHYLNAL
GIKARIREKLRKDGIDYVLHVEEYSSLLRFYELIGKNLQNEEKREKLEKVLSNHKGGN
FGLPLNFNAFKE
WASEYGVEFKTNGSQTIAIINDERISLGQWHTRNRVSKAVLVKMLRKLYEATKDEEV

KRMLHLIEGLEVVRHITTTNEPRTFYDLTVENYQNYLAGENGMIFVHNT

FIGURE 3

P)

NSILPEEWVPLIKNGKVKIFRIGDFVDGLMKANQGKVKKTGDTEVLEVAGIHAFSFD
RKSKKARVMAVKA
VIRHRYSGNVYRIVLNSGRKITITEGHSLFVYRNGDLVEATGEDVKIGDLLAVPRSVN
LPEKRERLNIVE
LLLNLSPEETEDIILTIPVKGRKNFFKGMLRTLRWIFGEEKRVRTASRYLRHLENLGYI
RLRKIGYDIID
KEGLEKYRTLYEKLVDVVRYNGNKREYLVEFNAVRDVISLMPEEELKEWRIGTRNG
FRMGTFVDIDEDFA
KLLGYYVSEGSARKWKNQTGGWSYTVRLYNENDEVLDDMEHLAKKFFGKVKRGK
NYVEIPKKMAYIIFES
LCGTLAENKRVPEVIFTSSKGVRWAFLEGYFIGDGDVHPSKRVRLSTKSELLVNGLV
LLLNSLGVSAIKL
GYDSGVYRVYVNEELKFTEYRKKKNVYHSHIVPKDILKETFGKVFQKNISYKKFREL
VENGKLDREKAKR
IEWLLNGDIVLDRVVEIKREYYDGYVYDLSVDEDENFLAGFGFLYAHNS

Q)

DSVTGETEIIIKRNGKVEFVAIEELFQRVDYRIGEKEYCVLEGVEALTLDNRGRLVWK
SV
PYVMRHRTNKRIYRVWFTNSWYLDVTEDHSLIGYMNTSKVKPGKPLKERLVEVKPG
ELGE
SVKSLITPNRAIAHGIRVNPIAVKLWELIGLLVGDGNWGGQSNWAKYNVGLSLGLDK
EEI
EEKILKPLKNTGIISNYYDKSKKGDVSILSKWLARFMVRYFKDESGSKRIPEFMFNLP
RE
YIEAFLRGLFSADGTVSLRKGVPEVRLTSVNPELSSSVRKLLWLVGVSNSMFVETNP
NRY
LGKESGTHSVHVRIKDKHRFAERIGFLLDRKATKLSENLGGHTSKKRAYKYDFDLVY
PKK
VEEIAYDGYVYDIEVEGTHRFFANGILVHNT

R)

KCLLPEEKVVLPEIGLVTLRELFELANEVVVKDEEKEVRKLGKMLTGVDERGNVKL
LNALYVWRVAHK
GEMIRVKVNGWYSVTVTPEHPFLTNRGWVKAGELKEGDYIAIPRRVYGNEDIMKFS
KIAKELGIKGDE
KEFYLAGASIDIPIKVLFLAPSKLVSAFLRGYFDAKGVVRENYIEVPLFEDLPLLILRFG
IVSRIEKS
TLKISGKRNLELFRKHVGFTDSEKAKALDELISKAKESERYPIIEELRRLGLLFGFTRN

ELRIEENPT
YEVIMEILERIERGSPNLAEKIAVLEGRIKEENYLRILEEEGLIENGKLTELGKELLEVW
RNREFDSK
DVDYVRNIVENLVFLPVEKVERIEYEGYVYDVTTETHNFVANGILVHNT

FIGURE 3

S)
QCFSGEEVIIVEKGKDRKVVKLREFVEDALKEPSGEGMDGDIKVTYKDLRGEDVRIL
TKDGFVKLLYVNK
REGKQKLRKIVNLDKDYWLAVTPDHKVFTSEGLKEAGEITEKDEIIRVPLVILDGPKI
ASTYGEDGKFDD
YIRWKKYYEKTGNGYKRAAKELNIKESTLRWWTQGAKPNSLKMIEELEKLNLLPLT
SEDSRLEKVAIILG
ALFSDGNIDRNFNTLSFISSERKAIERFVETLKELFGEFNYEIRDNHESLGKSILFRTWD
RRIIRFFVAL
GAPVGNKTKVKLELPWWIKLKPSLFLAFMDGLYSGDGSVPRFARYEEGIKFNGTFEI
AQLTDDVEKKLPF
FEEIAWYLSFFGIKAKVRVDKTGDKYKVRLIFSQSIDNVLNFLEFIPISLSPAKREKFLR
EVESYLAAVP
ESSLAGRIEELREHFNRIKKGERRSFIETWEVVNVTYNVTTETGNLLANGLFVKNS

T)
LCLTPDTYVVLGDGRIETIEDIVNAKERNVLSLDLDNLSIKIDTAIKFWKLRYNGNLSK
ITLSNNYELKA
TPDHCLLVLRDNQLKWIPAKDIKENDYIAMPFNYKVERKPISLLNLLKYLDITDVLIE
FDENSTIFEKIA
EYIRNNIKTSTKYKYLRNRRVPLKYLIEWNFDLDEIEKEAKYIYKSVAGTKKIPLFKL
DERFWYFAGLVL
GDGSIQDSKIRIAQTPLKDVKSILDETFPFLHNWISGNQVIISNPIIAEILEKLGMRNGKL
NGIIFSLPE
SYINALIAGYFDTDGCFSLLYDKKAKKHNLRMVLTSKRRDVLEKIGIYLNSIGILNTL
HKSREVYSLIIS
NKSLETFKEKIAKYLKIRKEAFINGYKTYKKEHEERFECDLLPVKEVFKKLTFEKGRK
EILKDSKIHIEN
WYKEKTNNIPREKLKTVLRYANNSEHKEFLEKIVNGDISFVRVKKVENIPYDGYVYD
LSIKHNQNFISNG
VISHNC

U)
KCLTGDTKVIANGQLFELRELVEKISGGKFGPTPVKGLKVIGIDEDGKLREFEVQYVY
KDKTERLIRIRT
RLGRELKVTPYHPLLVNRRNGEIKWVKAEELKPGDKLAVPRFLPIVTGEDPLAEWLG
YFLGGGYADSKEN
LIMFTNEDPLLRQRFMELTEKLFSDARIREITHENGTSKVYVNSKKALKLVNSLGNAH

IPKECWRGIRSF
LRAYFDCNGGVKGNAIVLATASKEMSQEIAYALAGFGIISRIQEYRVIISGSDNVKKFL
NEIGFINRNKL
EKALKLVKKDDPGHDGLEINYELISYVKDRLRLSFFNDKRSWSYREAKEISWELMKE
IYYRLDELEKLKE
SLSRGILIDWNEVAKRIEEVAEETGIRADELLEYIEGKRKLSFKDYIKIAKVLGIDVEHT
IEAMRVFARK
YSSYAEIGRRLGTWNSSVKTILESNAVNVEILERIRKIELELIEEILSDEKLKEGIAYLIF

FIGURE 3

U) cont.
LSQNELYWD
EITKVEELRGEFIIYDLHVPGYHNFIAGNMPTVVHNT

V)
SCVTGDTKVYTPDEREVKIRDFMNYFENGLIKEVSNRIGRDTVIAAVSFNSRIVGHPV
YRLTLESGRIIE
ATGDHMFLTPEGWKQTYDIKEGSEVLVKPTLEGTPYEPDPRVIIDIKEFYNFLEKIERE
HNLKPLKEAKT
FRELITKDKEKILRRALELRAEIENGLTKREAEILELISADTWIPRAELEKKARISRTRL
NQILQRLEKK
GYIERRIEGRKQFVRKIRNGKILRNAMDIKRILEEEFGIKISYTTVKKLLSGNVDGMAY
RILKEVKEKWL
VRYDDEKAGILARVVGFILGDGHLARNGRIWFNSSKEELEMLANDLRKLGLKPSEIIE
RDSSSEIQGRKV
KGRIYMLYVDNAAFHALLRFWKVEVGNKTKKGYTVPEWIKKGNLFVKREFLRGLF
GADGTKPCGKRYNFN
GIKLEIRAKKESLERTVEFLNDVADLLREFDVDSKITVSPTKEGFIIRLIVTPNDANYLN
FLTRVGYAYA
KDTYARLVGEYIRIKLAYKNIILPGIAEKAIELATVTNSTYAAKVLGVSRDFVVNRLK
GTQIGITRDFMT
FEEFMKERVLNGYVIEKVIKKEKLGYLDVYDVTCARDHSFISNGLVSHNC

W)
NCLTSNSKILTDDGYYIKLEKLKEKLDLHIKIYNTEEGEKSSNILFVSERYADEKIIRIK
TESGRVLEGS
KDHPVLTLNGYVPMGMLKEGDDVIVYPYEGVEYEEPSDEIILDEDDFAEYDKQIIKY
LKDRGLLPLRMDN
KNIGIIARLLGFAFGDGSIVKENGDRERLYVAFYGKRETLIKIREDLEKLGIKASRIYSR
KREVEIRNAY
GDEYTSLCEDNSIKITSKAFALFMHKLGMPIGKKTEQIYKIPEWIKKAPKWVKRNFLA
GLFGADGSRAVF
KNYTPLPINLTMSKSEELKENILEFLNEIKLLLAEFDIESMIYEIKSLDGRVSYRLAIVG
EESIKNFLGR
INYEYSGEKKVIGLLAYEYLRRKDIAKEIRKKCIKRAKELYKKGVTVSEMLKMDEFR
NEFISKRLIERAV

YENLDEDDVRISTKFPKFEEFIEKYGVIGGFVIDKIKEIEEISYDSKLYDVGIVSKEHNFI
ANSIVVHNC

X)
KCVDGDTLVLTKEFGLIKIKELYEKLDGKGRKIVEGNEEWTELEKPITVYGYKDGKI
VEI
KATHVYKGVSSGMVEIRTRTGRKIKVTPIHRLFTGRVTKDGLILKEVMAMHVKPGD
RIAV
VKKIDGGEYIKLDSSNVGEIKVPEILNEELAEFLGYLMANGTLKSGIIEIYCDDESLLER
VNSLSLKLFGVGGRIVQKVDGKALVIQSKPLVDVLRRLGVPEDKKVENWKVPRELL
LSPS

FIGURE 3

X) cont.
NVVRAFVNAYIKGKEEVEITLASEEGAYELSYLFAKLGIYVTISKSGEYYKVRVSRRG
NL
DTIPVEVNGMPKVLPYEDFRKFAKSIGLEEVAENHLQHIIFDEVIDVRYIPEPQEVYDV
T
TETHNFVGGNMPTLLHNT

1043968

Figure 4A

Intein B

MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVAND

cyclic insert (with flagg epitope)

IIVHNSIEQGQGGGMSMDYKDDDDKMRMLEGQAGGLITSGCIS

GDSLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESKVSRVFCT

Intein A

GKKLVYILKTRLGRTIKATANHRFLTIDGWKRLDELSLKEHIALPRK

LESSSLQLSIHGYH

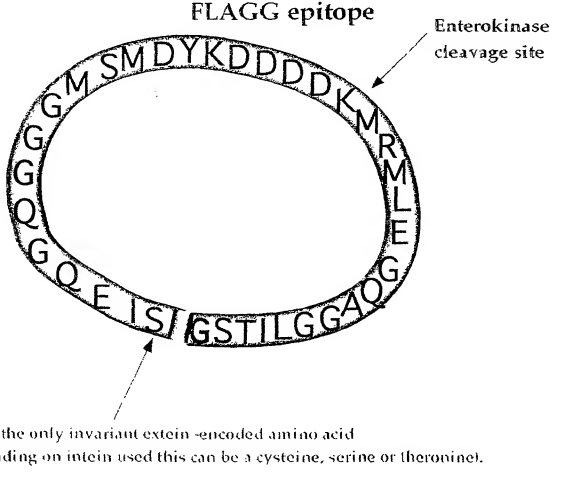


Figure 4B

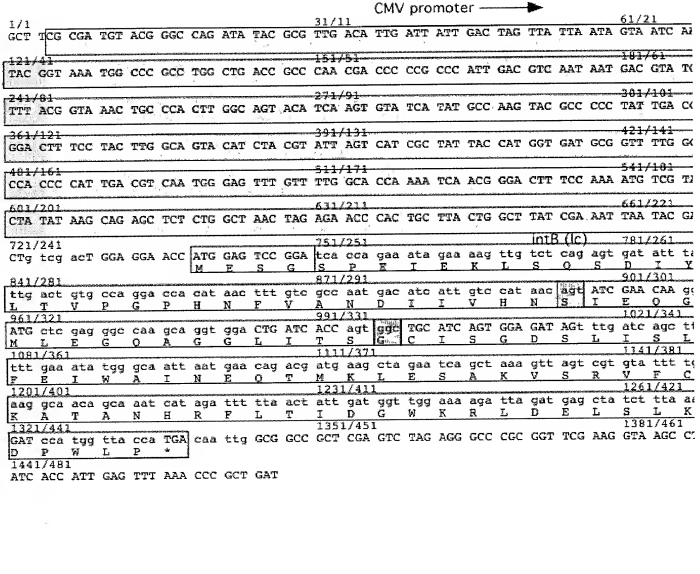


FIGURE 5

A)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AaAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

B)

MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGN
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

C)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AAAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAaATTTTTAAC
TATTGATGGTTGGAAAAGATT
AGAT
GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

D)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGN
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHKFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

E)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AaAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AgTTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

F)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGN
SLISLASTGKRVSIKDLLDEKDFEIWAVNEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

G)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGgCcGTGCCAGGGCCCCATAACTTTGTGGCCA
ATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AAAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT
GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

H)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLAVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGN
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

I)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTcCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AaAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

J)
MESGSPEIEKLSQSDIYWDSIVPITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGN
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

K)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AaAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAgGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT
GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

L)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGN
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKRLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

M)

ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AGAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAaATTTTTAAC
TATTGATGGTTGGAAAAGATT
AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

N)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGD
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHKFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

O)
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTcCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG

TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AGAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AATTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT
GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

P)
MESGSPEIEKLSQSDIYWDSIVPITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGD
SLISLASTGKRVSIKDLLDEKDFEIWAINEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

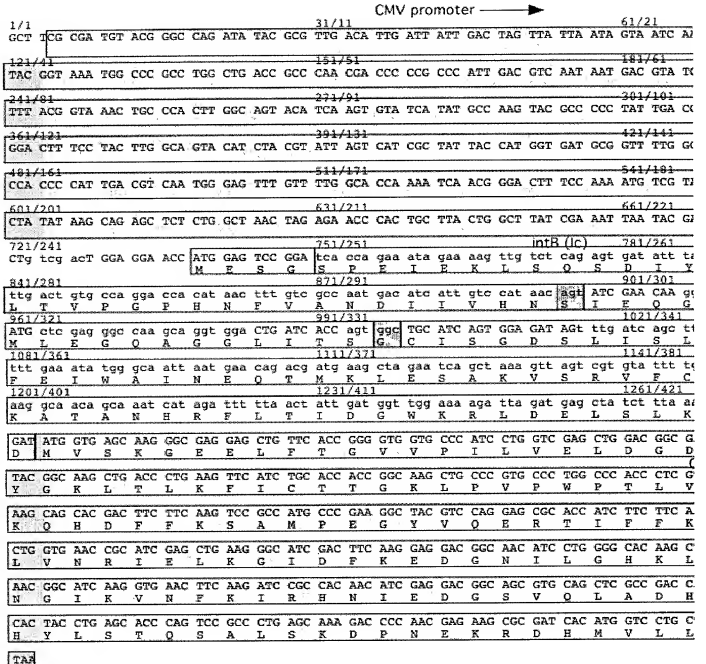
Q)

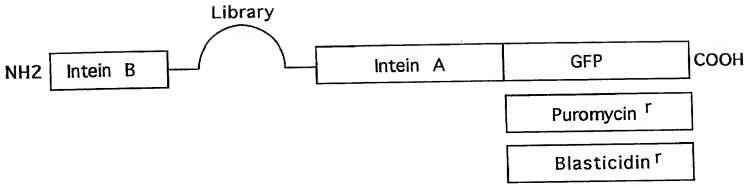
ATGGAGTCCGGATCACCAGAAATAGAAAAGTTGTCTCAGAGTGATATTTACTGG
GACTCCATCGTTTCTATTACGG
AGAC
TGGAGTCGAAGAGGTTTTTGATTTGACTGTGCCAGGGCCCCATAACTTTGTGGCC
AATGACATCATTGTCCATAAC
AGTG
AGGAGGACCTGGGATCCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCC
CCATCGGCGACGGCCCCGTGCT
GCTG
CCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAG
AAGCGCGATCACATGGTCCTGC
TGGA
GTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGGTC
GAACGGGGAATTCTCGCAGGTA
GACA
AGTCGATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGG
TCGAGCTGGACGGCGACGTAAA
CGGC
CACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTG
ACCCTGAAGTTCATCTGCACCA
CCGG
CAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAG
TGCTTCAGCCGCTACCCCGAC
CACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGA
CGGC
AACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGC
ATCGAGCTGAAGGGCATCGACT
TCAA
GGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAA
CGTCTATATCATGGCCGACAAG
CAGA
AGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACCTCGAGC
AAAAGCTGATATGCATCTCCGG
AGAT
AGTTTGATCAGCTTGGCGAGCACAGGAAAAAGAGTTTCTATTAAAGATTTGTTAG
ATGAAAAAGATTTTGAAATAT
GGGC
AGTTAATGAACAGACGATGAAGCTAGAATCAGCTAAAGTTAGTCGTGTATTTTGT
ACTGGCAAAAAGCTAGTTTAT
ATTT
TAAAAACTCGACTAGGTAGAACTATCAAGGCAACAGCAAATCATAGATTTTTAA
CTATTGATGGTTGGAAAAGATT
AGAT

GAGCTATCTTTAAAAGAGCATATTGCTCTACCCCGTAAACTAGAAAGCTCCTCTT
TACAATTAGGCCTCCGCGGCC
AGTA
CCCCTACGACGTCCCGGACTACGCTATCGATTAA

R)
MESGSPEIEKLSQSDIYWDSIVSITETGVEEVFDLTVPGPHNFVANDIIVHNSEEDLGSS
VQLADHYQQNTPIGDG
PVLL
PDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGSNGEFSQVDKSM
VSKGEELFTGVVPILVELDG
DVNG
HKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQ
HDFFKSAMPEGYVQERTIFF
KDDG
NYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGI
KVNFKIRHNIEDLEQKLIC
ISGD
SLISLASTGKRVSIKDLLDEKDFEIWAVNEQTMKLESAKVSRVFCTGKKLVYILKTRL
GRTIKATANHRFLTIDGW
KRLD
ELSLKEHIALPRKLESSSLQLGLRGQYPYDVPDYAIDZ

Figure 6




Library
NH2
Intein B
Intein A
GFP
COOH
Puromycin r
Blasticidin r

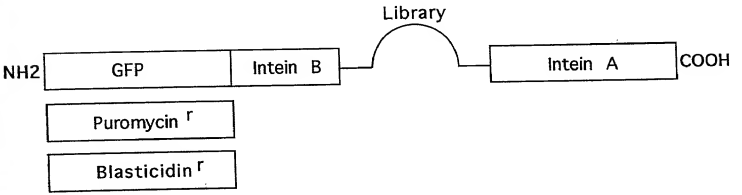

Library
NH2
GFP
Intein B
Intein A
COOH
Puromycin r
Blasticidin r

Figure 7

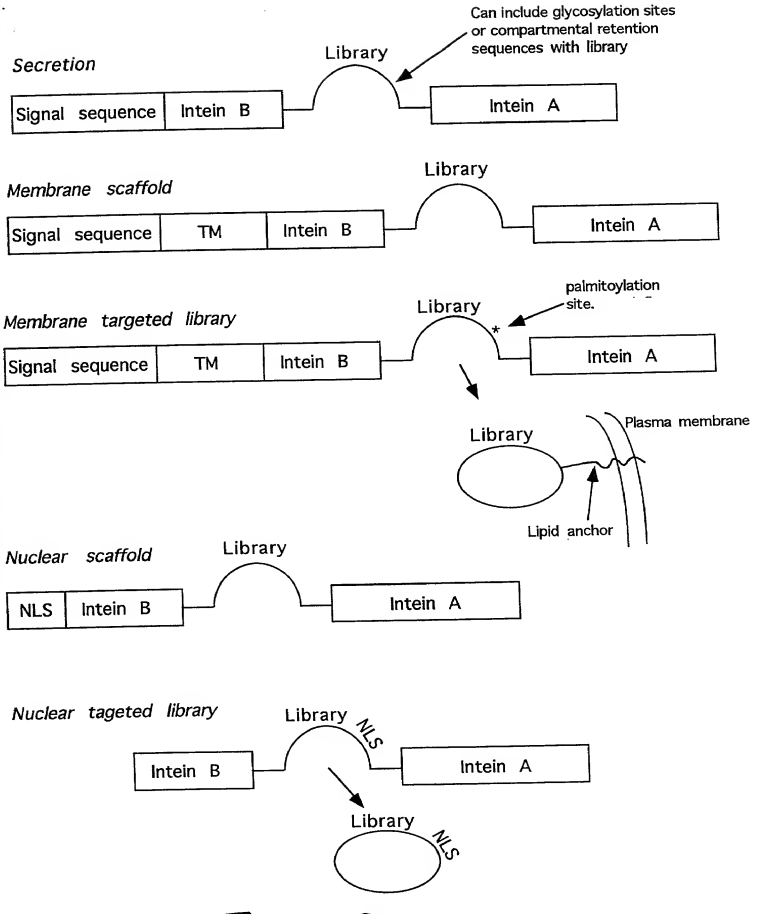

Secretion
Signal sequence
Intein B
Library
Can include glycosylation sites or compartmental retention sequences with library
Intein A
Membrane scaffold
Signal sequence
TM
Intein B
Library
Intein A
Membrane targeted library
Signal sequence
TM
Intein B
Library
palmitoylation site.
*
Intein A
Library
Plasma membrane
Lipid anchor
Nuclear scaffold
NLS
Intein B
Library
Intein A
Nuclear tageted library
Intein B
Library
NLS
Intein A
Library
NLS

Figure 8

1. Use a PCR mutagenesis or shuffling approach to mutate intein domains

2. Create a retroviral library of mutants

3. Infect cells and screen for those most effecient at cyclization (assayed indirectly by monitoring the release of IntA-TetRVP16 from its membrane location)

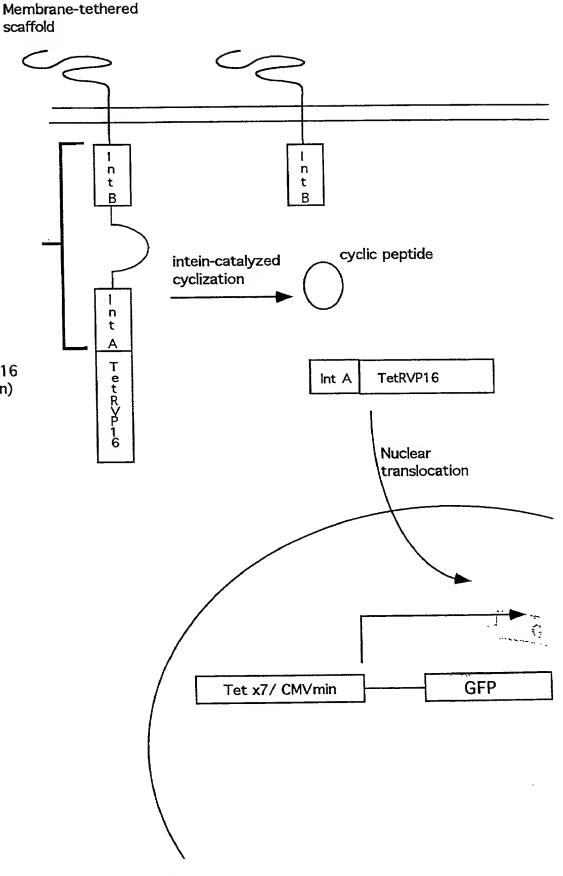


Figure 9

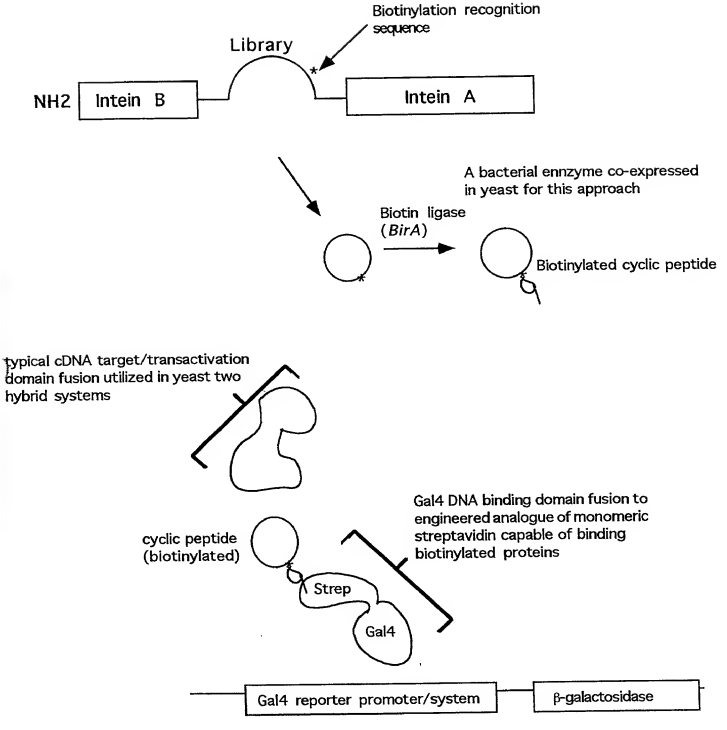

Biotinylation recognition sequence
Library
NH2
Intein B
Intein A
A bacterial ennzyme co-expressed in yeast for this approach
Biotin ligase (BirA)
Biotinylated cyclic peptide
typical cDNA target/transactivation domain fusion utilized in yeast two hybrid systems
Gal4 DNA binding domain fusion to engineered analogue of monomeric streptavidin capable of binding biotinylated proteins
cyclic peptide (biotinylated)
Strep
Gal4
Gal4 reporter promoter/system
β-galactosidase

Figure 10

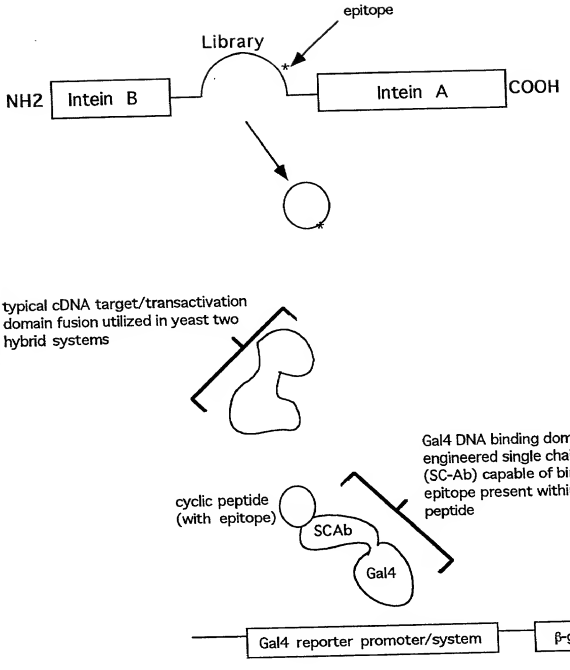
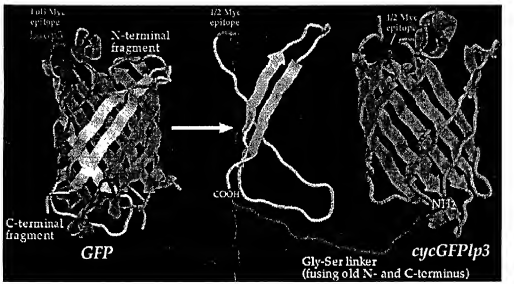
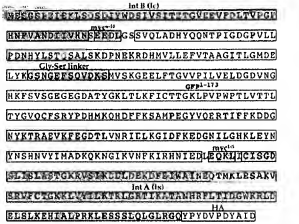

epitope
Library
NH2
Intein B
Intein A
COOH
typical cDNA target/transactivation domain fusion utilized in yeast two hybrid systems
Gal4 DNA binding domain fusion to engineered single chain antibody (SC-Ab) capable of binding to the epitope present within the cyclized peptide
cyclic peptide (with epitope)
SCAb
Gal4
Gal4 reporter promoter/system
β-galactosidase

Figure 11

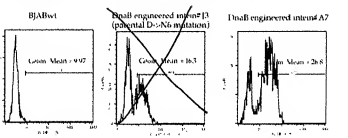
A.



B.



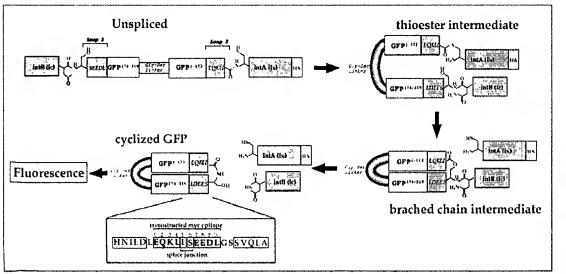
D.



E.

F.

C.



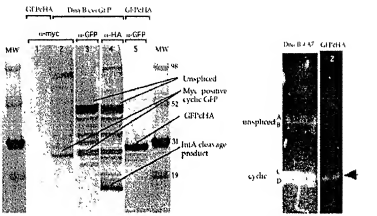


Figure 12

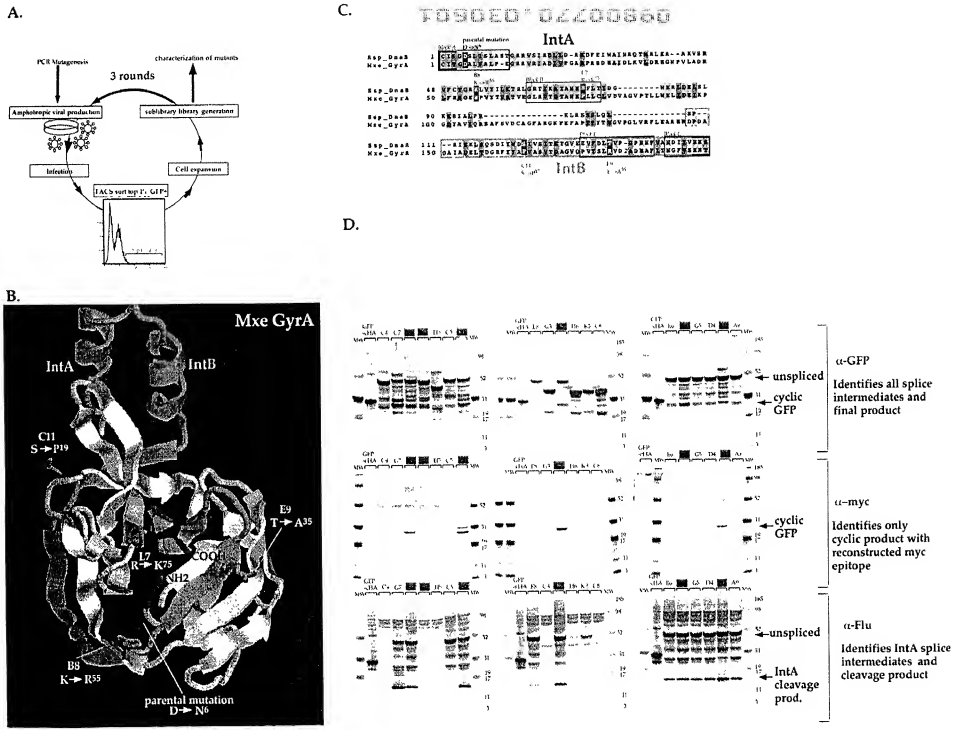
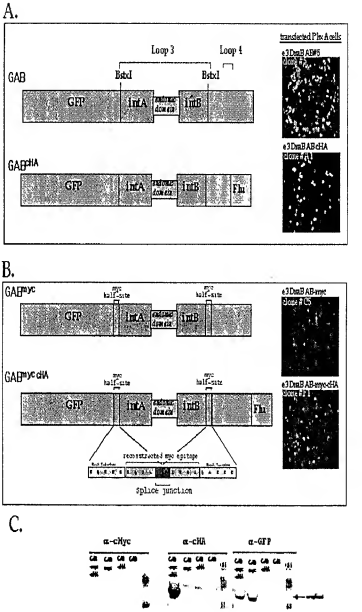

A.
PCR Mutagenesis
characterization of mutants
3 rounds
Amphotropic viral production
sublibrary library generation
Infection
Cell expansion
FACS sort top 1% GFP+
B.
Mxe GyrA
IntA
IntB
C11
S→P19
E9
T→A35
R→K75
COO-
NH2
B8
K→R55
parental mutation
D→N6
C.
IntA
IntB
D.
unspliced
cyclic
GFP
α-GFP
Identifies all splice intermediates and final product
cyclic
GFP
α-myc
Identifies only cyclic product with reconstructed myc epitope
unspliced
IntA
cleavage
prod.
α-Flu
Identifies IntA splice intermediates and cleavage product

Figure 13

Figure 14



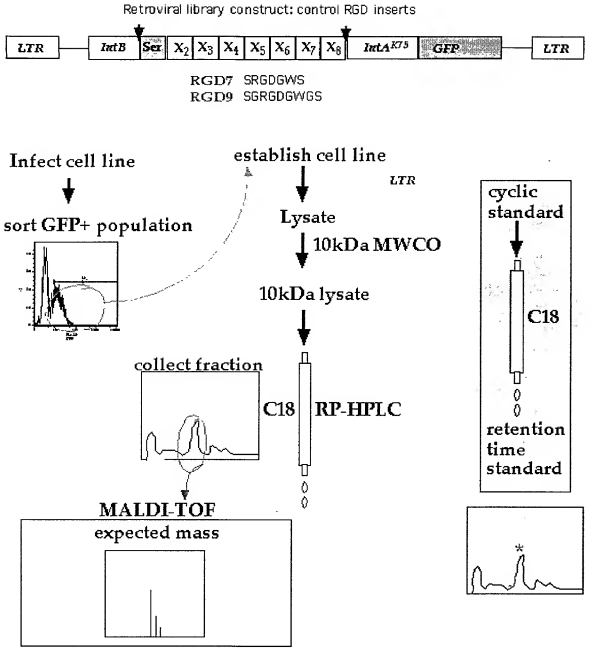

Retroviral library construct: control RGD inserts
LTR
IntB
Ser
X2 X3 X4 X5 X6 X7 X8
IntA K75
GFP
LTR
RGD7 SRGDGWS
RGD9 SGRGDGWGS
Infect cell line
sort GFP+ population
establish cell line
LTR
Lysate
10kDa MWCO
10kDa lysate
collect fraction
C18
RP-HPLC
MALDI-TOF
expected mass
cyclic standard
C18
retention time standard
*

Figure 15A

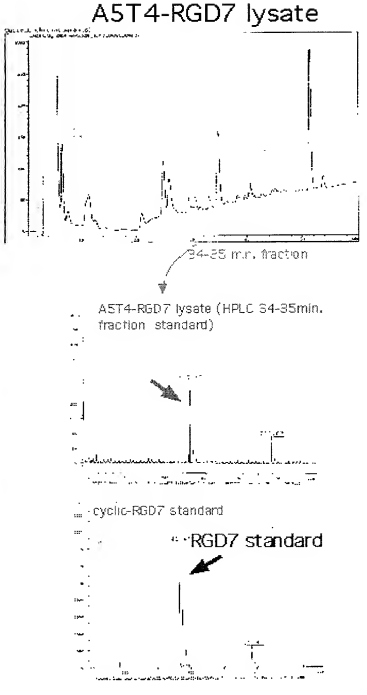

A5T4-RGD7 lysate
34-35 m.r. fraction
A5T4-RGD7 lysate (HPLC 34-35min. fraction standard)
cyclic-RGD7 standard
RGD7 standard

Figure 15 B

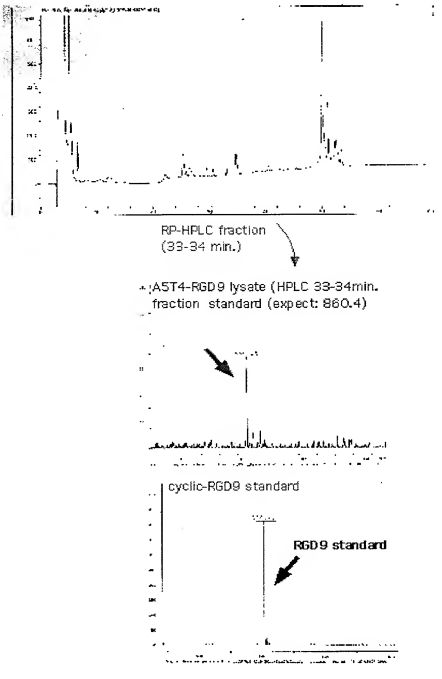
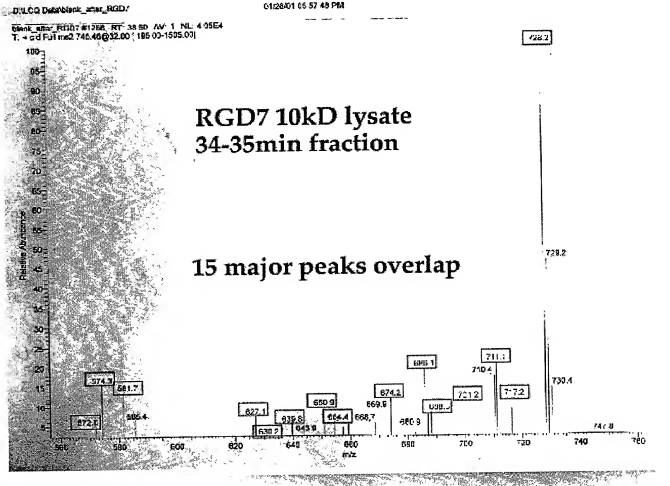

A5T4-RGD9 lysate
RP-HPLC fraction
(33-34 min.)
A5T4-RGD9 lysate (HPLC 33-34min.
fraction standard (expect: 860.4)
cyclic-RGD9 standard
RGD9 standard

Figure 15C

LC/MS fragmentation fingerprinting



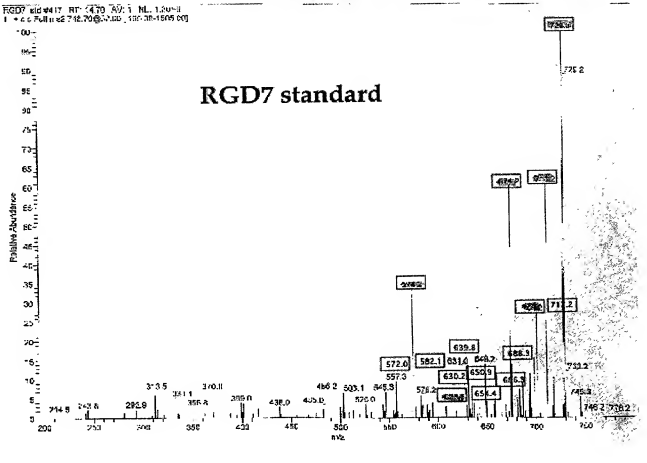


Figure 15D

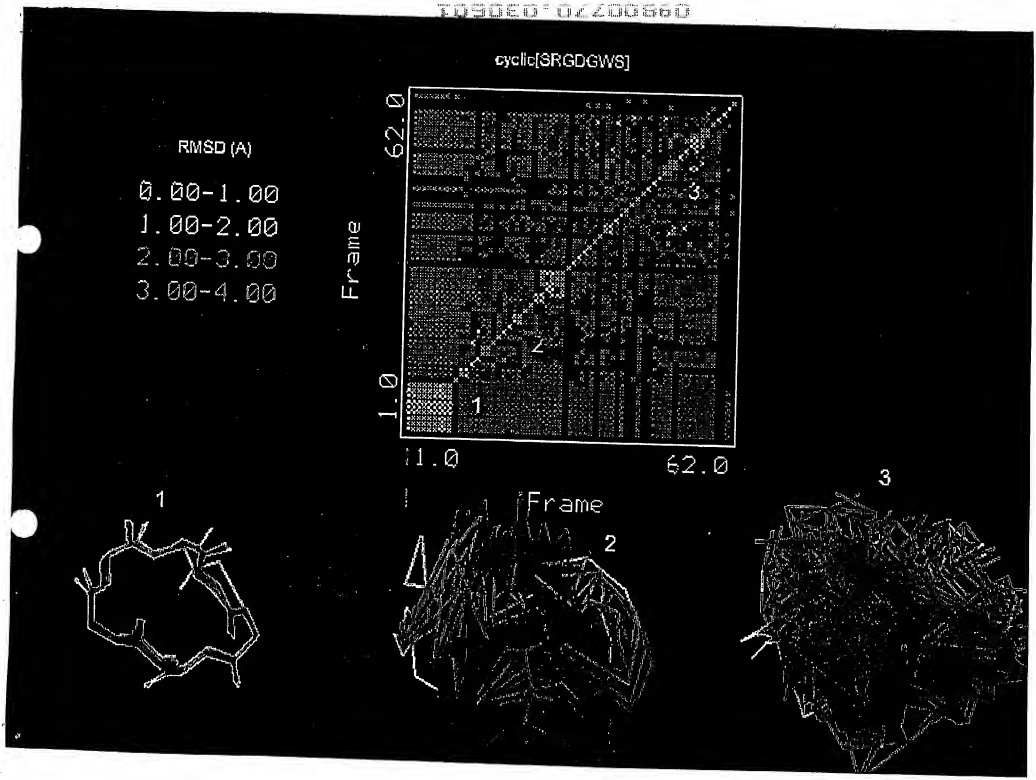


Figure 16

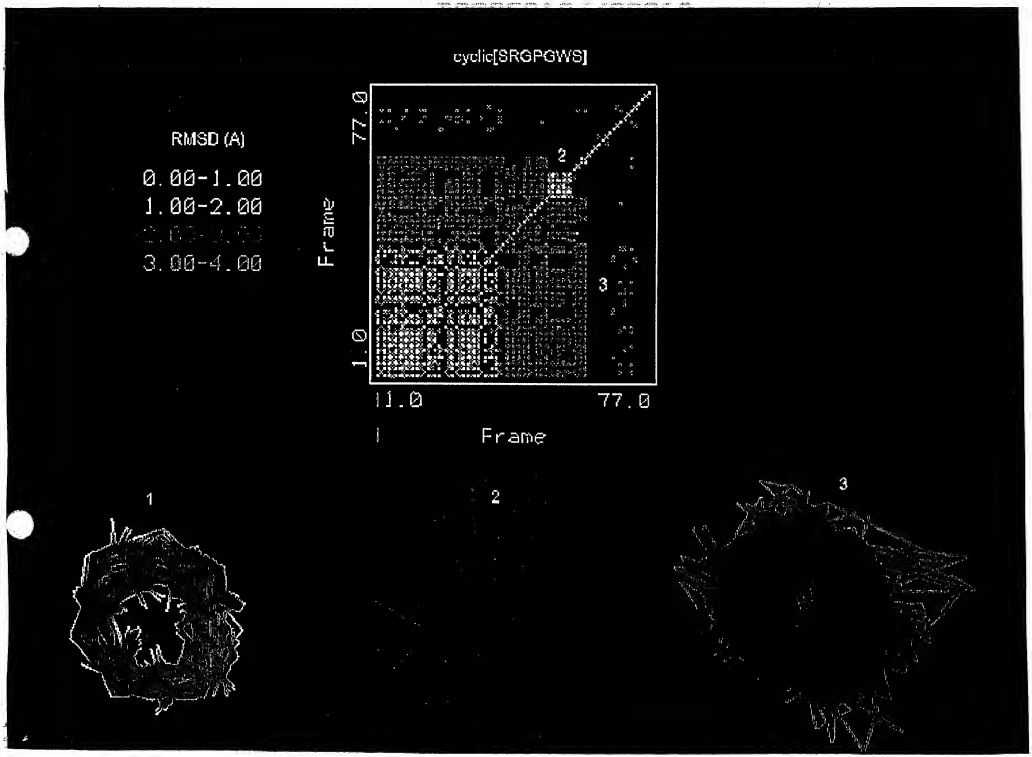

cyclic[SRGPGWS]
RMSD (A)
0.00-1.00
1.00-2.00
3.00-4.00
Frame
77.0
1.0
11.0
77.0
Frame
1
2
3
1
2
3

Figure 17